(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 December 2001 (06.12.2001)

PCT

(10) International Publication Number
WO 01/92291 A2

(51) International Patent Classification[7]: C07K

(21) International Application Number: PCT/IL01/00496

(22) International Filing Date: 30 May 2001 (30.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
136459 30 May 2000 (30.05.2000) IL

(71) Applicant *(for all designated States except US)*: **GALIM - GALIL IMMUNOLOGY LTD.** [IL/IL]; Migdal Bldg., S. Industrial Zone, Kiryat Shmona, P.O. Box 90000, 12100 Rosh Pina (IL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GROSS, Gideon** [IL/IL]; Moshav Almagor, 12 340 M.P. Korazim (IL). **SCHLESINGER, Yaaqov** [IL/IL]; Kibbutz Yiftah, 13840 M.P. Upper Galilee (IL).

(74) Agent: **BEN-AMI, Paulina**; Webb, Ben-Ami & Associates, P.O. Box 2189, 76121 Rehovot (IL).

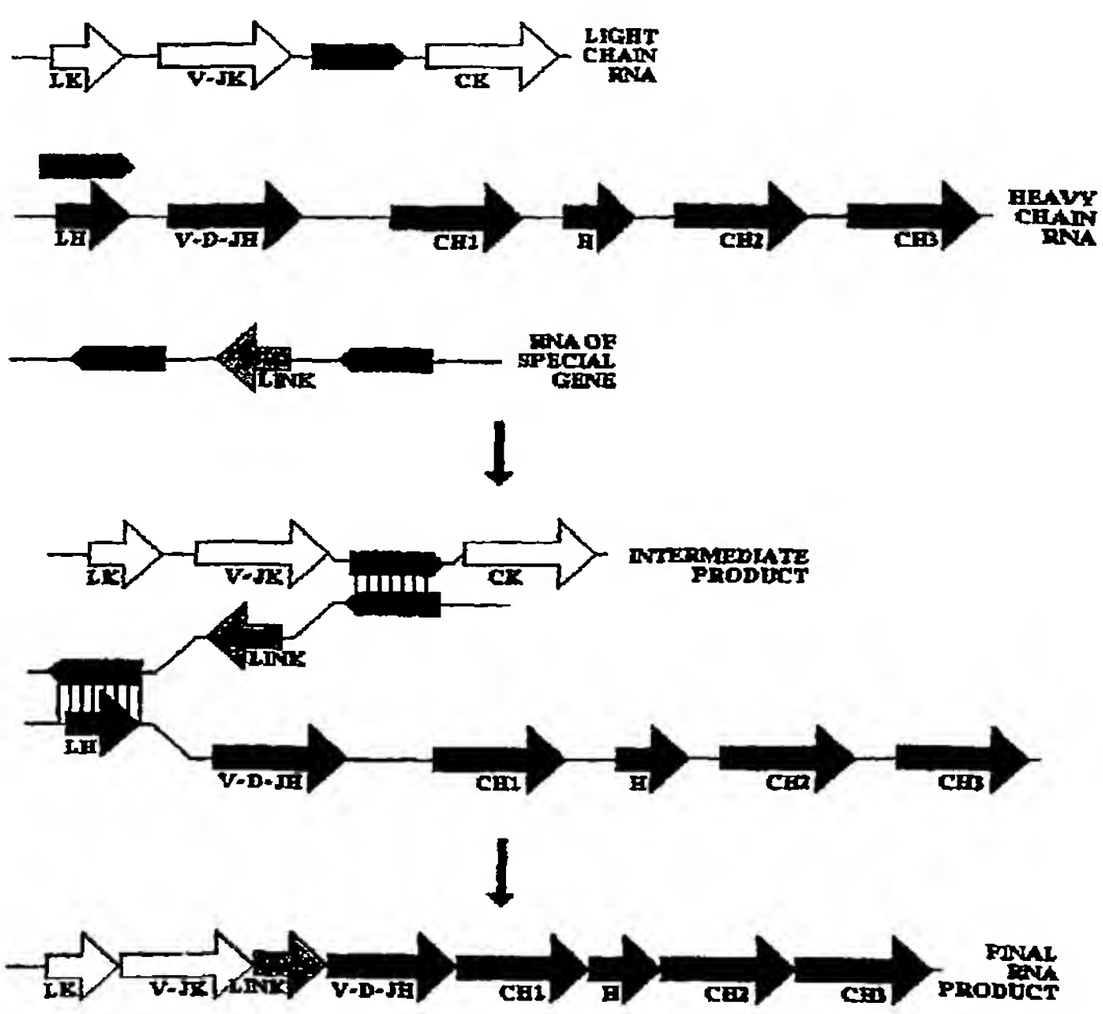
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: ANTIBODY AND T-CELL RECEPTOR LIBRARIES



(57) Abstract: The invention relates to RNA trans-splicing-mediated covalent intracellular fusion of transcripts of two different genes A and B in a cell, particularly in an immune cell, in which said genes A and B are expressible, and to DNA constructs encoding RNAs useful in such processes. The genes A and B may encode the light and heavy chains of an antibody or the α and β chains of a T-cell receptor. This process leads to a unique and novel approach for creating immune antibody and T-cell receptor libraries.

WO 01/92291 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

ANTIBODY AND T-CELL RECEPTOR LIBRARIES

FIELD OF THE INVENTION

The present invention relates to RNA trans-splicing-mediated covalent intracellular fusion of transcripts of two different genes A and B in a cell, particularly in an immune cell, in which said genes A and B are expressible, and to DNA constructs encoding RNAs useful in such processes. The genes A and B may encode the light and heavy chains of an antibody or the α and β chains of a T-cell receptor. This process leads to a unique and novel approach for creating immune antibody and T-cell receptor libraries.

GLOSSARY/ABBREVIATIONS

Acceptor splice site - a consensus sequence at the 3' end of an intron, which is recognized and cleaved during splicing.

Branch point - an intronic sequence near the 3' end of an intron, which forms a lariat structure with the 5' end of the same intron during splicing.

BP – branch point.

Donor splice site - a consensus sequence at the 5' end of an intron, which is recognized and cleaved during splicing

Fab fragment – a proteolytic fragment of the immunoglobulin molecule resulting from digestion with papain and comprising the VH-CH1 domain of the heavy chain and the VL-CL1 domain of the light chain, joined by a single-interchain disulfide bond.

fTS – "facilitated" trans-splicing, as defined in the specification.

Fv fragment – Ig fragment comprising only the VL and the VH domains or TCR fragment comprising only the Vα and Vβ domains.

Hinge (H) region – an extended proline-rich peptide sequence between the CH1 and CH2 domains of the heavy chain that has no homology with the other domains and imparts flexibility to the antibody molecule.

Ig – immunoglobulin.

Leader peptide – a short sequence of amino acids at the amino terminus of a newly synthesized protein which allows the growing polypeptide to cross the endoplasmic reticulum membrane and is cleaved thereafter.

mAb – monoclonal antibody.

mRNA - messenger RNA.

PCR – polymerase chain reaction.

pre-mRNA - a primary transcript before it is fully processed into mRNA.

Primary transcript - a newly synthesized RNA molecule comprising introns and exons.

Reverse transcriptase (RT) - RNA-directed DNA polymerase, a viral polymerase that catalyzes in a host cell the synthesis of a DNA strand complementary to the viral RNA.

RT-PCR - reverse transcriptase-polymerase chain reaction.

ScFv - single-chain Fv: Ig fragment comprising only the VL and the VH domains connected by a peptide linker or TCR fragment comprising only the Vα and Vβ domains connected by a peptide linker.

Sense strand - the non-template strand of a duplex DNA in a gene, which is identical in sequence to the RNA transcript.

Signal peptide – see **Leader peptide**

snRNAs - small nuclear RNAs: specialized RNA molecules involved in splicing reactions.

snRNPs - small nuclear ribonucleoproteins: specialized RNA-protein complexes involved in splicing reactions.

Spliceosome – the splicing apparatus which consists of several snRNPs and a number of additional individual proteins, referred to as splicing factors.

Splicing - process by which introns are removed from primary transcripts and the exons are joined to form a contiguous sequence encoding a functional polypeptide.

TCR - T-cell receptor.

Template - the strand of a duplex DNA from which RNA complementary to said strand is synthesized.

TS – trans-splicing; **sTS** – single trans-splicing; **dTS** – double trans-splicing.

Variable heavy domain (VH) – the portion of the heavy chain that comprises part of the variable region of an antibody molecule.

Variable light domain (VL) – the portion of the light chain that comprises part of the variable region of an antibody molecule.

BACKGROUND OF THE INVENTION

Many, if not most, eukaryotic genes have nucleotide sequences containing one or more intervening sequences of DNA that do not code for the amino acid sequence of the polypeptide product. These non-translated inserts interrupt the otherwise precisely colinear relationship between the nucleotide sequence of the gene and the amino acid sequence of the polypeptide it encodes. Such intervening non-translated segments in genes are called introns and the coding segments are called exons.

The expression of the genetic information contained in a segment of DNA involves the generation of a molecule of RNA. In a process called transcription, an enzyme system converts the genetic information of a segment of DNA into an RNA strand with a base sequence complementary to one of the DNA strands. The DNA strand that serves as template for RNA synthesis is called the template strand.

A newly synthesized RNA molecule is called a primary transcript. A primary transcript for a eukaryotic messenger RNA (mRNA) typically contains sequences encompassing one gene. The sequences encoding the polypeptide, however, usually are not contiguous. As in the DNA molecule from which it is synthesized, an RNA molecule contains introns and exons. In a process called splicing, the introns are removed from the primary transcript and the exons are joined (spliced) to form the contiguous sequence of a mature mRNA specifying a functional polypeptide. The splice points between exons are typically determined by consensus sequences that act as signals for the splicing genes.

A phenomenon called RNA trans-splicing (hereinafter RNA TS), namely bimolecular joining of exons from different RNA species, has been well documented in trypanosomes, nematodes, trematodes and plant mitochondria and

chloroplasts [reviewed in (Huang and Hirsh, 1992). RNA trans-splicing has also been suggested to be responsible for fused RNA transcripts observed in a variety of mammalian cells such as those of double-Ig isotype expression in human B cells and in mouse B cells transgenic for human Ig heavy chain genes (Fujieda et al., 1996; Shimizu et al. 1989). In addition, the ability of nuclear extracts from mammalian cells to facilitate TS has been clearly demonstrated in vitro, using special RNA substrates (Solnick, 1985; Konarska et al., 1985).

The experimental scheme of RNA trans-splicing in vitro between two model mRNAs **1** and **2** containing two exons and one intron according to these two publications is illustrated in **Fig. 1** herein. According to this scheme, together with the normal splicing of each transcript resulting in the deletion of the intron and joining of the two exons of the same transcript (cis-splicing), there are also obtained transcripts in which exons of different mRNAs are joined by deletion of the introns via an intermediate product formed by hybridization of the two transcripts **1** and **2** through their intronic sequences.

Trans-splicing reactions not requiring RNA-RNA interactions among the substrates have been observed in vitro (Chiara and Reed, 1995) and in vivo (Caudevilla et al., 1998). Most importantly, base-pairing-promoted TS has recently been demonstrated in a number of studies to occur inside the nucleus of living mammalian cells. Spliceosome-mediated single TS (sTS) was evaluated as a gene therapy tool (Puttaraju et al., 1999; also see review by Garcia-Blanco et al., 2000). Ribozyme-mediated TS has also been studied by a number of laboratories as a potential therapeutic tool (reviewed by Phylactou et al, 1998; Watanabe and Sullenger, 2000).

Antibodies and T cell receptors are homologous families of proteins that share structural homology and have similar functions, namely, to confer specificity in antigen recognition. The specificity of the immune response in vertebrates is dictated by a very large repertoire of these molecules.

Antibodies or immunoglobulins (Ig) are antigen-binding proteins present on the B-cell membrane and secreted by plasma cells. Antibodies are Y-shaped

proteins consisting of four polypeptide chains – two heavy (H) and two light (L) chains, interconnected by disulfide bridges.

Each heavy and light chain in an Ig molecule contains an amino-terminal variable (V) region – V_H and V_L - that consists of 100-110 amino acids, differs greatly from one Ig to the other and comprises the binding sites that are complementary to specific structural features of the antigen molecule, thus permitting recognition and binding of the antigen by the antibody. There are two distinct families of light chains, called kappa (κ) and lambda (λ), which differ somewhat in the sequences of their constant regions. In humans, 60% of the light chains are κ and 40% are λ, whereas in mice 95% of the light chains are κ and only 5% are λ. A single antibody molecule contains either κ or λ light chains but never both. There are 3 subtypes of λ light chains in mice (λ1, λ2, λ3) and 4 subtypes in humans (λ1, λ2, λ3, λ4), distinguished by minor differences in their amino acid sequences.

The remainder of the molecule of each heavy and light chain in an Ig molecule is virtually constant (C) within a class of Ig - C_H and C_L. There are 5 subclasses of heavy chain constant (C_H) regions, designated μ, δ, γ, ε, and α, and characterized by different amino acid patterns. The heavy chains of a given molecule determine the class of that antibody. In humans, the chain composition of the 5 Ig classes is as follows: IgG (γ), IgA (α), IgM(μ), IgD(δ) and IgE(ε). Some heavy chains (δ, γ, and α) also contain a proline-rich hinge region.

In addition, both light and heavy chains of an Ig molecule contain several homologous units of about 110 amino acid residues. Within each unit, designated a domain, an intrachain disulfide bond forms a loop of about 60 amino acids. Light chains contain one variable domain (V_L) and one constant domain (C_L); heavy chains contain one variable domain (V_H) and either 3 or 4 constant domains (C_H1, C_H2, C_H3 , and C_H4), depending on the antibody class.

Monoclonal antibodies (mAbs), generated by a population of identical cells grown in a cell culture, are homogeneous and recognize the same specific part of the antigen. Since their introduction to science in the mid 70's, mAbs generated

against antigens of choice have made a tremendous impact in numerous fields: in medicine - as effective reagents for neutralization of pathogens, blocking of undesirable molecular or cellular interactions or mimicking favorable ones; in medical, environmental and forensic diagnostics - as specific and sensitive tools for the detection and quantification of a whole universe of substances; in biological and biochemical research – in the analysis and isolation of molecules located inside, on the surface or in the environment of living cells; in industry - as effective reagents for material purification, product analysis, biosensor development and many other applications,

The usefulness of mAbs depends on their specificity, namely, their ability to discriminate between closely-related structures, and their affinity, namely, the strength of their interaction with the antigen. Successful isolation of optimal mAbs requires a large starting repertoire of antibody-expressing entities, from which those encoding specific, high affinity products can be easily selected.

Three main technologies for generating mAbs are in common use, each with significant drawbacks alongside its obvious advantages:

1. Immunization of laboratory mice or any other rodent with the relevant antigen and isolation of splenocytes, including antibody-producing B cells, which are then immortalized by fusion with myeloma cells to produce B cell hybridomas (Harlow and Lane, 1988). Hybridomas preserve the ability of B cells to synthesize antigen-specific antibodies, and large amounts of products are easily obtained. This technology frequently yields high affinity antibodies, which are generated and selected in vivo by the affinity maturation process in the course of the immune response, prior to the isolation of the B cells. However, making hybridomas is a slow and tedious procedure. In addition, a successful fusion between B cells and the myeloma partner which stably retains all relevant chromosomes, is a rare event. Therefore, the products represent only a minor fraction of the actual antigen-specific repertoire, with frequent loss of the exact fine specificity requested. Human hybridomas are technically and ethically hard (but not impossible) to obtain. As an alternative, lines of transgenic mice harboring sizable portions of the human immunoglobulin heavy and light chain gene loci have been generated at several

laboratories, thus permitting the production of murine B-cell hybridomas secreting fully human mAbs (reviewed by Bruggemann and Neuberger, 1996).

2. Generation of 'quasi' immune antibody libraries expressed primarily on filamentous phages, also known as 'phage display libraries' (Winter et al., 1994), or on ribosomes, yeast and other vehicles. The raw material for these libraries is mRNA of B cells derived from immune humans or laboratory animals. The pools of VH and VL genes are amplified by RT-PCR separately, each with a specific set of primers. The resulting genes encoding the VH and VL chains are then shuffled randomly and cloned in vectors, which drive their expression on phages either as scFv or as Fab fragments. In this way, large repertoires of antibodies are formed and displayed on the surface of the filamentous phage. Phage display is easy, fast, and generates huge libraries, which allow isolation of antigen-specific phage-antibodies in a minimal number of simple enrichment steps. These libraries are indeed enriched for antigen-binding clones, but these are usually only of low-to-moderate affinity, as reconstitution in this random manner of authentic VH and VL chain pairs from the same affinity mature B cells, or of an otherwise effective pair of chains, is an event of low probability.

3. By-passing immunization in two ways (Winter et al., 1994): first, RT-PCR amplification of mRNA pools of both chains from non-immune (naïve) individuals (or animals), followed by random shuffling of H and L chain cDNA products; second, generation of semi-synthetic repertoires, by introducing all possible nucleotide permutations into a certain coding stretch(es), usually the third complementarity-determining region (CDR3) of the H chain [for example (Nissim et al., 1994). Both these strategies only rarely result in antibodies displaying high affinities typical to those generated in vivo in the secondary immune response. As a result, most antibodies isolated from such non-immune libraries require further improvement, which is exercised on each clone separately and is lengthy and laborious.

A basic expression modality widely used in antibody display libraries are antibody scFv fragments, as they make up the smallest stable functional units which retain the binding properties of antibodies, encoded by only one gene.

Reconstitution of the antibody Fv structure and binding capacity is made possible by a usually glycine- and serine-rich flexible peptide linker, which joins the carboxy terminus of either VH or VL with the amino terminus of the other V domain. Most linkers used for this purpose are of 15-18 amino acids and they are added genetically as short segment between the VL and VH coding sequences in the scFv gene (Whitlow and Filpula, 1992).

Diabodies are formed from two interacting scFv fragments, which are constructed with short peptide linkers (5-10 amino acids) or even as VL and VH domains joined directly, with no linker, so that VL of one scFv is forced to pair with the VH of the other, creating two antigen binding sites (Holliger et al., 1993). Diabodies designed to form two identical binding sites for the same antigen, are referred to as bivalent diabodies, whereas those designed to form two distinct binding sites, directed at two different antigens, are termed bi-specific diabodies. Diabodies can serve as the basic binding unit for antibody library construction (McGuinness et al., 1996).

A general approach, which enables the faithful reconstitution of the full high affinity antigen-specific antibody repertoire from immune humans or animals, through expression on phages or another suitable genetic package, is still lacking. The key to such an achievement is the development of a technique for the covalent joining of the rearranged VH and VL genes, or gene products, from the same B cell, performed simultaneously on large B cell populations, followed by isolation of the joined products in an expressible form. An interesting attempt in this direction has been reported (Embleton et al., 1992), suggesting a procedure for in-cell PCR to link both genes. However, this protocol poses a number of severe technical difficulties and it failed to be generalized for wider application.

T-cell receptors (TCR) are antigen-binding proteins present on the T-cell membrane. Unlike the membrane-bound antibody on B cells, the T-cell receptor recognizes an antigen only in the form of a short peptide displayed together with MHC molecules on the surface of antigen-presenting cells. TCRs are heterodimeric transmembrane proteins made of α and β chains linked by an intrachain disulfide bond. Amino acid sequencing of the $\alpha\beta$ heterodimers shows a domain structure

very similar to that of immunoglobulins; thus, they are members of the immunoglobulin (Ig) superfamily. The α and β chains, like those of the Ig heavy and light chains, have constant and variable regions. The amino-terminal domain in both chains exhibits marked sequence variation, but the sequences of the remainder of each chain are conserved. Thus, the TCR domains – one variable (V) and one constant (C) – are structurally homologous to the V and C domains of Igs, and the TCR molecule is thought to resemble an Fab fragment.

In spite of antibody and TCR overall structural similarities, functional scFv TCRs are much harder to obtain, although a number of successful attempts have been reported (e.g. Ward, 1992). One could envisage that the ability to isolate TCRs with a predefined specificity from immune individuals may have several important applications. Such TCR genes can serve as the basis for specific reagents in adoptive immunotherapy of cancer, infectious diseases or other disorders (with or without further in-vitro improvement), or as immunogens, designed to eliminate harmful T cells, for example, those involved in autoimmune diseases. Such TCRs can also be invaluable in studies aimed at characterizing the exact profile of immune responses. So far, the production of useful TCR display libraries has been hampered by a number of technical difficulties, one of which is the randomness associated with gene shuffling, as described above for antibody libraries.

International PCT Publication No. WO 92/19619 describes methods for producing a library of DNA molecules capable of expressing a fusion polypeptide on the surface of a filamentous phage particle (via phagemids) and producing heterodimeric receptors such as antibodies and T-cell receptors.

US Patent No. 5,969,108 describes libraries of DNA encoding respective chains of multimeric specific binding pair members such as the VH and VL chains of an antibody, in which said binding pair members are displayed in functional form at the surface of a secreted recombinant genetic display package containing DNA encoding said binding pair member or a polypeptide component thereof, by virtue of the specific binding pair member or a polypeptide thereof being expressed as a fusion with a capsid component of the recombinant genetic display package. The antibody members are thus obtained with the different chains thereof

expressed, one fused to the capsid component and the other in free form for association with the fusion partner polypeptide. Packaging in a phagemid as an expression vector produces antibody libraries said to have a much greater diversity in the antibody VL and VH chains than by conventional methods.

US Patents Nos. 5,498,531 and 5,780,272 describe in vitro intron-mediated combinatorial methods for generating a variegated population of ribonucleic acids encoding chimeric gene products comprising admixing a variegated set of splicing constructs under trans-splicing conditions. The method can be used for generating diverse antibody libraries.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide reagents and methods for the manipulation of nucleic acids.

The present invention is based on a novel strategy via trans-splicing (TS) for covalent intracellular joining of pairs of genes that are normally expressed and associated in native cells such as the heavy and light chains of an antibody or the α and β chains of a T-cell receptor that normally are expressed and associated in immune B and T cells, respectively.

For this purpose, the present invention provides a DNA construct consisting of a transcription unit useful for covalent intracellular joining of selected exons from transcripts of two different genes A and B in a cell in which said genes A and B are expressible, said transcription unit comprising promoter/enhancer elements and a template for RNA synthesis, wherein said template encodes an RNA transcript of the general formula:

RCA - SP - RCB

wherein

RCA represents a nucleotide segment having one or more sequences, each reverse complementary to one or more sequences of pre-mRNA of gene A and genes related thereto, said sequences of pre-mRNA of gene A and genes related thereto being situated downstream to an exon selected to be spliced to pre-mRNA of gene B or genes related thereto;

SP is either a spacer sequence or a sequence of the formula:

SP1 - LEX - SP2

wherein

SP1 represents a spacer sequence;

LEX represents an exon, said exon encoding a flexible peptide linker or a part thereof preceded by branch point and acceptor splice sequences and followed by a donor splice sequence; and

SP2 represents a spacer sequence; and

RCB represents a nucleotide segment having one or more sequences, each reverse complementary to one or more sequences of pre-mRNA of gene B and genes related thereto, said sequences of pre-mRNA of gene B and genes related thereto being situated upstream to an exon in pre-mRNA of gene B or genes related thereto selected to be spliced to said selected exon of gene A.

The cells in which the pair of genes A and B are normally expressed and associated in some form for their functional activity are preferably immune cells such as B and T lymphocytes. In one embodiment, the immune cell is a T lymphocyte and either gene A encodes a T-cell receptor (TCR) α chain and gene B encodes a TCR β chain, or gene A encodes a TCR β chain and gene B encodes a TCR α chain. In one most preferred embodiment, the immune cell is an antibody-producing B lymphocyte and either gene A encodes an antibody light (L) chain and gene B encodes an antibody heavy (H) chain, or gene A encodes an antibody heavy (H) chain and gene B encodes an antibody light (L) chain.

The invention further provides transgenic animals, preferably mice, harboring the DNA constructs useful for covalent intracellular joining of selected exons from transcripts of an antibody L and H chains.

The invention still further provides cDNA libraries useful for the production of gene libraries encoding scFv fragments of antibodies or TCRs of interest, said gene libraries, expression vectors comprising said gene libraries and phage-display libraries comprising a plurality of recombinant phages having said vectors.

DESCRIPTION OF THE FIGURES

Fig. 1 depicts a model scheme for in vitro TS between two separate mRNAs, termed Gene 1 (first row) and Gene 2 (2nd row), mediated by base-paring, as described by Solnick, 1985 and Konarska et al., 1985. In each experiment, an intron (black arrow between Exon 1 and Exon 2) in one of the two RNA substrates was designed to harbor a stretch reverse-complementary to an intronic sequence in the second RNA substrate. The two RNA species were pre-hybridized (Intermediate Product) and then subjected to the splicing machinery of human HeLa cells nuclear extract, resulting in trans-splicing efficiency (bottom, right) which was up to 30% of that observed for cis-splicing (bottom, left).

Fig. 2 depicts a schematic representation of the dTS gene of the invention (third row, RNA of Special Gene) and the predicted scFv-encoding RNA product of arrangement VL-spacer-linker-spacer-VH (last row, Final RNA Product). First row - Light Chain (Vκ) RNA (exons - white arrows; intron - black arrow); Second row - Heavy Chain (VH) RNA (exons - gray arrows; intron - black arrow); Third row - RNA of Special Gene according to the invention. Upon annealing of this dTS-mediating transcript to the Vκ and VH pre-mRNAs via the appropriate reverse-complementary sequences, two TS events occur, which lead to the formation of the scFv transcript (last row - Final RNA Product): (a) TS between the donor splice site at the 3' end of the V-Jκ exon and the acceptor splice site at the 5' end of the linker exon (Intermediate Product, white arrows); and (b) TS between the donor splice site at the 3' end of the linker exon and the acceptor splice site at the 5' end of the V-D-JH exon (Intermediate Product, gray arrows).

LH - VH leader; Lκ - Vκ leader; V-D-JH - rearranged VH gene; V-Jκ - rearranged Vκ gene; C - constant chain; H (alone) - hinge region; CH - constant-heavy; Cκ - constant κ; LINK - linker.

Fig. 3 depicts the restriction map of clone 965-14 described in Example 1, obtained by cloning into expression vector pBJ1-Neo two fragments of the L and H chains of the monoclonal antibody (anti-trinitrophenyl mAb) produced by hybridoma SP6. This clone is designed to mediate sTS between the donor splice site at the 3' end of the linker exon (shown here), and the acceptor splice site at the 5' end of SP6 VH exon (not shown), upon annealing of the reverse-complementary sequence (underlined with an arrow) to the LH exon.

Xh - XhoI; B - BamHI; N - NotI; a - acceptor splice site; d - donor splice site; L - leader exon of SP6 Vκ; Vκ - SP6 Vk.

Fig. 4 demonstrates single TS in SP6 hybridoma cells (transfectant 965-2) transfected with clone 965-14, as described in Example 2. Upper line - shows a sequence of a part of clone 965-14, showing from left to right the 3' end of the Vκ-linker exon (upper case) and the donor splice site (lower case). Middle line - shows a sequence of a part of the genomic SP6 VH gene, showing from left to right the acceptor splice site at the 3' end of the SP6 leader-VH intron (lower case) and the 5' end of the VH exon (upper case); Lower line - shows the sense strand of cDNA produced from mRNA of 965-2 cells, demonstrating accurate joining of both coding sequences from upper and middle lines. This DNA sequence is derived from the PCR products in clones 979-1 and 979-3, and is identical in both.

Fig. 5 is a schematic representation of clones 75-2 and 1033-1 described in Examples 3a and 3c, respectively, and their expected TS product. Clone 1033-1 is designed to mediate sTS between the donor splice site at the 3' end of its SP6 Vκ-linker exon, and the acceptor splice site at the 5' end of the SP6 VH exon in clone 75-2, upon annealing of the reverse-complementary sequences (marked by vertical lines).

Fig. 6 is a schematic representation of clones 1083-1 and 1026-2 described in Examples 3b and 3d, respectively, and their expected TS product. Clone 1026-2 is designed to mediate sTS between the acceptor splice site at the 5' end of its linker-

SP6 VH exon and the donor splice site at the 3' end of the SP6 Vκ exon in clone 1083-1, upon annealing of the reverse complementary sequences (marked by vertical lines).

B - BamHI; H - HindIII; N - NotI; Xb - XbaI; R - EcoRI; li - linker; L - SP6 VH leader; VH - SP6 VH; CH - human IgG H chain constant domain; h - hinge;Vk - SP6 Vκ; a - acceptor splice site; d - donor splice site.

Figs. 7A-7B show a scheme of the transcript clone 1034-2 (**7A**), described in Example 3e and designed to mediate dTS of SP6 Vκ to VH, and the DNA sequence thereof (**7B**).

In **Fig. 7A**, black boxes represent reverse-complementary sequences to SP6 genes. The Xh-N fragment is reverse complementary to the 3' end of the Jκ-Cκ intron. The Xb-R fragment is reverse complementary to SP6 VH leader exon-intron junction. The N-B fragment is the first spacer, which is the intron in the human Cγ1 gene, between CH1 and the hinge exons. The B-Xb fragment is the second spacer, which comprises a part of the mouse JH3-JH4 intron. Restriction sites refer to those in 1034-2 DNA.

B - BamHI; N - NotI; R - EcoRI; Xb - XbaI; Xh - XhoI; a - acceptor splice site; d - donor splice site.

Fig. 7B shows the DNA sequence of SP6 dTS construct 1034-2. Important restriction sites are underlined. Bases 1-132 - a 120 bp XhoI/NotI fragment with a sense strand reverse-complementary to the 3' region of the mouse Jκ-Cκ intron (positions 4493-4613 in GenBank Accession V00777). Bases 133-516 - a fragment of the human Cγ1- hinge intron (positions 514-896 in GenBank Accession Z17370), containing a branch point (BP) and an acceptor splice site in its 3' end. Bases 517-562 - the peptide linker coding sequence. Bases 563-712 - a spacer region from the mouse JH3-JH4 intron (nucleotides 1181-1330 in GenBank Accession V00777), containing a donor splice site in its 5' end. Bases 713-744 - the sequence reverse-complementary to the SP6 VH transcript.

Figs. 8A-C show the sequences of TS products obtained following COS7 transfections, as described in Example 5.

Fig. 8A. Single TS in COS7 cells transfected with clones 75-2 + 1033-1 described in Examples 3a and 3c, respectively. Upper line - shows a sequence of a part of clone 1033-1, showing from left to right the 3' end of the linker exon (upper case) and the donor splice site (lower case). Middle line - shows a sequence of a part of the SP6 VH gene in clone 75-2, showing from left to right the acceptor splice site at the 3' end of the SP6 leader-VH intron (lower case) and the 5' end of the VH exon (upper case); Lower line - shows the sense strand of cDNA produced from mRNA of transfected COS7 cells, demonstrating accurate joining of both coding sequences from upper and middle lines.

Fig. 8B. Single TS in COS7 cells transfected with clones 1083-1 + 1026-2 described in Examples 3b and 3d, respectively. Upper line - shows a sequence of a part of SP6 κ chain gene from clone 1083-1, showing from left to right the 3' end of the V-Jκ exon (upper case) and the donor splice site (lower case). Middle line - shows a sequence of a part of clone 1026-2, showing from left to right the acceptor splice site immediately upstream to the linker-SP6 VH exon (lower case) and the 5' end of that VH exon (upper case); Lower line - shows the sense strand of cDNA produced from mRNA of transfected COS7 cells, demonstrating accurate joining of both coding sequences from upper and middle lines.

Fig. 8C. Facilitated TS in COS7 cells transfected with clones 75-2 + 1083-1 +1034-2, described in Examples 3a, 3b and 3e, respectively. Upper line - shows a sequence of a part of SP6 κ chain gene from clone 1083-1, showing from left to right the 3' end of the V-Jκ exon (upper case) and the donor splice site (lower case). Middle line - shows a sequence of a part of the SP6 VH gene in clone 75-2, showing from left to right the acceptor splice site at the 3' end of the SP6 leader-VH intron (lower case) and the 5' end of the VH exon (upper case). Lower line - shows the sense strand of cDNA produced from mRNA of transfected COS7 cells, demonstrating in-frame, accurate splicing of SP6 Vκ exon to the VH exon.

Fig. 9 shows the 19 sequences from mouse VH leader-intron junctions, used as annealing targets in the design of clone 1068-2, as described in Example 9. The sequences, in upper case, are numbered 1-19. The GT dinucleotides at the donor splice sites are underlined. Under each sequence: Left, gene annotations according to the ABG database. Right, GenBank accession of the corresponding gene. Note that some sequences are present in more than one functional mouse VH germline gene. 42 mouse VH genes are represented.

Fig. 10 shows the 11 synthetic oligonucleotides that were used for the assembly of the XbaI/EcoRI fragment of clone 1068-2 (Example 9). The first five oligonucleotides comprise the "upper" strand in the synthetic double stranded fragment (see **Fig. 11**). Oligonucleotides 6-11 comprise the "lower" strand. "P" at the 5' end of an oligonucleotide indicates the phosphate group that was chemically added to the 5' end of that oligonucleotide. These phosphate groups are necessary for the ligation of the oligonucleotides following their hybridization.

Fig. 11 shows the sequence of the double-stranded synthetic XbaI-EcoRI DNA fragment in clone 1068-2 (see Fig. 12), encoding the 19 segments reverse complementary to the mouse VH sequences presented in Fig. 9. Restriction sites incorporated into the sequence are underlined.

Figs. 12A-12B show a scheme of the transcript of the mouse dTS construct, clone 1068-2 (Example 9) and the DNA sequence thereof, respectively.

In **Fig. 12A,** black boxes represent reverse-complementary sequences to mouse κ and H chain genes. The Xh-N fragment is reverse-complementary to the 3' end of the Jκ-Cκ intron. The Xb-R fragment is encoded by the synthetic DNA fragment in Fig. 11, harboring the 19 segments which cover all mouse VH leader exon-intron junctions. The N-B fragment is the first spacer, which is the intron in the human Cγ1 gene, between CH1 and the hinge exons. The B-Xb fragment is the second spacer, which comprises a part of the mouse JH3-JH4 intron. Restriction sites correspond to those in 1034-2 DNA.

B - BamHI; N - NotI; R - EcoRI; Xb - XbaI; Xh - XhoI; a - acceptor splice site; d - donor splice site.

Fig. 12B shows the DNA sequence of the mouse dTS construct 1068-2. Important restriction sites are underlined. Bases 1-132: A 120 bp XhoI/NotI fragment with a sense strand reverse-complementary to the 3' region of the mouse Jκ-Cκ intron (positions 4493-4613 in GenBank Accession V00777). Bases 133-516 - a fragment of the human Cγ1- hinge intron (positions 514-896 in GenBank Accession Z17370), containing a BP and an acceptor splice site in its 3' end. Bases 517-562 - the peptide linker coding sequence. Bases 563-712 - a spacer region from the mouse JH3-JH4 intron (nucleotides 1181-1330 in GenBank Accession V00777), containing a donor splice site in its 5' end. Bases 713-1137 - the synthetic sequence containing the regions reverse-complementary to all mouse VH germ-line genes (see Fig. 11).

Fig. 13 shows the 24 sequences from human VH leader-intron junctions, used as annealing targets in the design of the human dTS construct (Example 10). Sequences, in upper case, are numbered. The GT dinucleotides at the donor splice sites are underlined. Under each sequence: Left, gene annotations according to the ImMunoGeneTics database. Right, GenBank accession of the corresponding gene. Note that some sequences are present in more than one functional human VH germline gene. 39 human VH genes are represented.

Fig. 14 shows the sequence of the double-stranded synthetic XbaI-EcoRI DNA fragment, encoding the 24 transcript segments reverse-complementary to the human VH sequences presented in Fig. 13. Restriction sites incorporated into the sequence are underlined.

Fig. 15 shows the sequences of the 9 synthetic oligonucleotides that were used for the synthesis of the fragment targeting TS of human L chain transcripts. The first 4 oligonucleotides comprise the "upper" strand in the synthetic double-stranded fragment (see Fig. 16). Oligonucleotides 5-9 comprise the "lower" strand. "P" at

the 5' end of an oligonucleotide indicates the phosphate group that was chemically added to the 5' end of that oligonucleotide. These phosphate groups are necessary for the ligation of the oligonucleotides following their hybridization.

Fig. 16 shows the sequence of the double stranded synthetic DNA fragment allowing RNA base-pairing with JL-CL intron junctions of all functional human L chain genes. Restriction sites incorporated into the sequence are underlined.

Figs. 17A-17B show a scheme of the transcript of the human dTS construct and its DNA sequence, respectively.

In **Fig. 17A,** black boxes represent reverse-complementary sequences to human L and H chain genes. The Xh-N fragment is is encoded by the synthetic DNA fragment in Fig. 16, harboring the 3 segments which cover the 3' end of the JL-CL introns of the functional human L chain genes. The Xb-R fragment is encoded by the synthetic DNA fragment in Fig. 14, harboring the 24 segments which cover all human VH leader exon-intron junctions. The N-B fragment is the first spacer, which is the intron in the human Cγ1 gene, between CH1 and the hinge exons. The B-Xb fragment is the second spacer, which comprises a part of the mouse JH3-JH4 intron. B - BamHI; N - NotI; R - EcoRI; Xb - XbaI; Xh - XhoI; a - acceptor splice site; d - donor splice site.

Fig. 17 B shows the DNA sequence of the human dTS construct. Important restriction sites are underlined. Bases 1-316 - the sequence containing the regions reverse-complementary to the human J-C κ and λ introns (see Fig. 16). Bases 317-700 - a fragment of the human Cγ1- hinge intron (positions 514-896 in GenBank Accession Z17370), containing a BP and an acceptor splice site in its 3' end. Bases 701-746 - the peptide linker coding sequence. Bases 746-896 - a spacer region from the mouse JH3-JH4 intron (nucleotides 1181-1330 in GenBank Accession V00777), containing a donor splice site in its 5' end. Bases 897-1438 - The sequence containing the regions reverse-complementary to all human VH germ-line genes (see Fig. 14).

Fig. 18 is a schematic representation of the dTS gene of the invention for generating TCR scFv libraries, and the predicted scFv-encoding RNA product attached to TCR β chain constant region gene. Upon annealing of the dTS-mediating transcript to the Vα and Vβ pre-mRNAs via the appropriate reverse-complementary sequences, two TS events, which lead to the formation of the scFv transcript, occur: (a) TS between the donor splice site at the 3' end of the V-Jα exon and the acceptor splice site at the 5' end of the linker exon; and (b) TS between the donor splice site at the 3' end of the linker exon and the acceptor splice site at the 5' end of the V-D-Jβ exon.

Lα: Vα leader; Lβ: Vβ leader; V-D-Jβ: rearranged Vβ gene; V-Jα: rearranged Vα gene; C: constant; Cα: constant α; Cβ: constant β; LINK: linker.

DETAILED DESCRIPTION OF THE INVENTION

Some terms used in the present application are defined as follows:

"facilitated TS" (fTS) - when SP is a spacer sequence, the RNA encoded by the construct of the invention will facilitate TS between the two transcripts A and B, a single trans-splicing (sTS) process herein referred to as "facilitated TS" (fTS).

"double TS" (dTS) - when SP is a sequence of the formula: SP1 - LEX - SP2, the RNA encoded by the construct of the invention will directly participate in the splicing reactions, a process herein referred to as "double TS" (dTS), since it involves two sTS events.

"antibodies of interest" and "TCRs of interest" refer to antibodies and TCRs, respectively, elicited by the immune system against an antigen/immunogen such as, but not being limited to, a whole cell (bacterium or virus), a protein, a peptide or a non-proteinaceous molecule coupled to a protein carrier.

"scFv" refers to single-chain Ig or TCR fragments comprising only the VL+VH or Vα+Vβ domains, respectively, connected by a peptide linker. When the peptide linker is long enough, usually of 15 amino acids or more, the VL and VH chains from the same polypeptide can functionally associate and create a (potential) antigen-binding site. When the peptide linker is too short, usually of less than 10

amino acids, the physical constraint prevents association of the VL and VH from the same polypeptide. Rather, VL from one polypeptide can then pair with a VH from a second polypeptide, thus creating a (potential) bivalent antigen-binding structure, referred to as a diabody. The term scFv herein refers to fragments when the VL and VH or Vα and Vβ chains are linked both by short or longer peptides.

"immune individual" refers to a human that has been exposed to, and exhibits, a T-cell response against an antigen/immunogen of interest.

The present invention provides a DNA construct consisting of a transcription unit useful for covalent intracellular joining of selected exons from transcripts of two different genes A and B in a cell in which said genes A and B are expressible, said transcription unit comprising promoter/enhancer elements and a template for RNA synthesis, wherein said template encodes an RNA transcript of the general formula:

RCA - SP - RCB

wherein

RCA represents a nucleotide segment having one or more sequences, each reverse-complementary to one or more sequences of pre-mRNA of gene A and genes related thereto, said sequences of pre-mRNA of gene A and genes related thereto being situated downstream to an exon selected to be spliced to pre-mRNA of gene B or genes related thereto;

SP is either a spacer sequence or a sequence of the formula:

SP1 - LEX - SP2

wherein

SP1 represents a spacer sequence;

LEX represents an exon, said exon encoding a flexible peptide linker or a part thereof preceded by branch point and acceptor splice sequences and followed by a donor splice sequence; and

SP2 represents a spacer sequence; and

RCB represents a nucleotide segment having one or more sequences, each reverse-complementary to one or more sequences of pre-mRNA of gene B and genes related thereto, said sequences of pre-mRNA of gene B and genes related

thereto being situated upstream to an exon in pre-mRNA of gene B or genes related thereto selected to be spliced to said selected exon of gene A.

The cells in which the DNA construct of the invention will cause the covalent joining of exons from transcripts of the genes A and B are cells in which this pair of genes A and B are normally expressed and associated in some form for their functional activity, and are preferably immune cells such as B and T cells.

In one embodiment, the immune cell is a human T lymphocyte and either gene A encodes a human T-cell receptor (TCR) α chain and gene B encodes a human TCR β chain, or gene A encodes a human TCR β chain and gene B encodes a human TCR α chain.

In one most preferred embodiment, the immune cell is an antibody-producing B lymphocyte and either gene A encodes an antibody light (L) chain and gene B encodes an antibody heavy (H) chain, or gene A encodes an antibody heavy (H) chain and gene B encodes an antibody light (L) chain.

The H and L chains are preferably the H and L chains of any suitable vertebrate antibody such as avian and, preferably, a mammalian antibody such as, but not being limited to, a murine, e.g. mouse, and, most preferably, a human antibody.

In one embodiment, the invention relates to a novel strategy for covalent intracellular joining of antibody heavy (H) and light (L) chain transcripts in antibody-producing B cells, via a special genetic construct designed to induce RNA trans-splicing (TS) of the two genes. The RNA encoded by this construct of the invention will either facilitate TS between the two transcripts by fTS or will directly participate in the splicing reactions by dTS, as described above. The resulting mRNA products of fTS will code for the variable (V) regions of the L and H chains (VL and VH) linked by a short peptide linker, while those resulting from dTS will code for the variable (V) regions of the L and H chains (VL and VH) linked by a longer flexible peptide linker. The same events of fTS and dTS occur with TCR Vα and Vβ chains in T lymphocytes.

Trans-spliced mRNA obtained either by fTS or dTS is then ready for single-step reverse transcriptase-polymerase chain reaction (RT-PCR) cloning, and

expression as either diabodies or monovalent scFv fragments, respectively, and prepared for display via suitable genetic packages such as bacterial cells, bacterial spores, yeast cells, ribosomes or, preferably, bacteriophages, e.g. filamentous phages.

The scheme according to the present invention provides a universal solution for the yet unsolved problem of intracellular functional joining of antibody H and L chain genes in antibody-producing cells. This accomplishment is mandatory for the generation of immune antibody libraries faithfully reconstituting B-cell immune repertoires, as easily-accessible, highly-enriched sources for high affinity monoclonal antibodies (mAbs) specific for antigens of choice. Such a powerful tool should be widely applicable for the production of invaluable reagents in medicine, diagnostics, research and industry.

The basic concept of the present invention was herein evaluated through the design and construction of a model experimental system with regard to antibody genes, in well-defined in-cell systems. The results obtained herein with these systems demonstrate the basic molecular processes required for VL and VH intracellular joining. To the best of our knowledge, it is shown here for the first time: (a) directed TS of RNA transcribed from two different chromosomal genes in mammalian cells, and (b) accurate facilitated trans-splicing (fTS) of transcripts of two different genes, which produces an in-frame scFv-encoding mRNA. It is further described herein the rationale and assembly of genetic constructs designed to promote fTS or dTS in diverse populations of immune cells, so as to allow the production of immune libraries from different mammalian species, with emphasis on human and mouse.

The idea underlying the present invention is that TS can be exploited intracellularly for the covalent fusion of antibody H and L chain or TCR α and β chain transcripts via a third RNA molecule - the product of a specially-engineered genetic construct, introduced into, and properly expressed by, B or T cells, respectively.

In order to better understand the basic concept of the invention, reference is made to the Background section of the present application, wherein the structure of

the immunoglobulin molecules is described, and to the following description regarding multigene organization of Ig genes, rearrangement of the variable and heavy chain gene segments and expression of Ig genes.

For each of the three types of polypeptide chains of an Ig molecule - heavy, κ and λ chains, diversity in the variable regions is generated by a similar mechanism. The genes for these polypeptides are encoded by separate multigene families situated on different chromosomes. In germ-line DNA, each of these multigene families contains several coding sequences, called gene segments, separated by non-coding regions. During B-cell maturation, these gene segments (one version of each segment) are rearranged and brought together to create a complete functional immunoglobulin gene.

The κ and λ light-chain families contain V (variable), J (joining) and C (constant) gene segments; the rearranged VJ segments encode the variable region of the light chains. The heavy-chain family contains V, D (for diversity), J and C gene segments; the rearranged VDJ segments encode the variable region of the heavy-chain. The C gene segments encode the constant regions. Each V gene segment is preceded at its 5' end by a small exon that encodes a short signal or leader (L) peptide that guides the heavy or light chain through the endoplasmic reticulum. The signal peptide is cleaved from the nascent light and heavy chains before assembly of the mature immunoglobulin molecule. Thus, the amino acids encoded by this leader sequence do not appear in the immunoglobulin molecule.

Expression of both κ and λ light chain genes requires rearrangement of the variable-region V and J gene segments. Rearranged κ and λ genes contain the following regions in order from the 5' to 3' end: a short leader (L) exon, a non-coding sequence (intron), a joined VJ gene segment, a second intron, and a C gene segment. Upstream from each leader gene segment is a promoter sequence. The rearranged light-chain sequence is transcribed by RNA polymerase from upstream to the L exon through the C segment and downstream to the stop codon (including 5' and 3' non-translated sequences), generating a light-chain primary RNA transcript. Following capping, splicing and polyadenylation, the resulting light-chain mRNA then exits from the nucleus, binds to ribosomes and is translated into

the light-chain protein. The leader sequence at the N-terminus pulls the growing polypeptide chain into the lumen of the rough endoplasmic reticulum and is then cleaved, as explained above.

Generation of a functional Ig heavy-chain gene requires two separate rearrangement events within the variable region. A D_H gene segment first joins to a J_H segment; a V_H segment then moves next to and joins the resulting D_HJ_H segment to generate a $V_HD_HJ_H$ unit that encodes the entire variable region. In heavy-chain DNA, rearrangement produces a rearranged gene consisting of the following sequences starting from the 5' end: a short L exon, an intron, a joined VDJ segment, another intron and a series of C_H gene segments. As with the light-chain genes, a promoter sequence is located a short distance upstream from each heavy-chain leader sequence.

Once heavy-chain gene rearrangement is accomplished, RNA polymerase can bind to the promoter sequence and transcribe the entire heavy-chain gene, including the introns. Initially, both C_μ and C_δ gene segments are transcribed. Differential polyadenylation and RNA splicing remove the introns and process the primary transcript to generate mRNA, encoding either C_μ or C_δ. These two mRNAs then are translated, and the leader peptide of the resulting nascent polypeptide is cleaved, generating mature μ and δ chains. Since two different heavy-chain mRNAs are produced following heavy-chain variable-region gene rearrangement, a mature, immunocompetent B cell expresses both IgM and IgD with identical antigenic specificity on its surface.

The phenomenon of allelic exclusion ensures that both for the heavy and the light chains, only one allele of the two, which are present on the homologous chromosomes, is functional. This guarantees that all antibodies produced by each B cell bear the same variable region and, hence, an identical specificity. As with the Ig genes, rearrangement of the TCR β-chain genes (and less stringent for the TCR α-chain genes) exhibits allelic exclusion.

In the process of splicing, some sequences play a special role: (a) a consensus sequence at the 5' end of an intron, which is recognized and cleaved

during splicing, termed "donor splice site"; (b) a consensus sequence at the 3' end of an intron, which is recognized and cleaved during splicing, termed "acceptor splice site"; and (c) an intronic sequence near the 3' end of an intron, which forms a lariat structure with the 5' end of the same intron during splicing, termed "branch point" (BP).

In addition, for the regulation of Ig gene transcription, regulatory sequences are required - promoters, situated upstream from the transcription initiation site, that promote initiation of RNA transcription in a specific direction, and enhancers, situated in the J-C introns and downstream to the C region genes, that activate transcription from the promoter sequence in an orientation-independent manner.

In one embodiment of the present invention, genes A and B encode an antibody H and L chains. In one preferred embodiment, as depicted in **Fig. 2** for dTS, gene A encodes a L chain and gene B encodes a H chain and the arrangement of the dTS product will be VL-linker-VH. In another embodiment, A encodes a H chain and B encodes a L chain and the arrangement of the dTS product will be VH-linker-VL.

In another embodiment of the present invention, genes A and B encode a TCR α and β chains. In one preferred embodiment, as depicted in **Fig. 18,** A encodes an α chain and B encodes a β chain and the arrangement of the dTS product will be Vα-linker-Vβ. In another embodiment, A encodes a β chain and B encodes an α chain and the arrangement of the dTS product will be Vβ-linker-Vα.

In the embodiment of the invention wherein the dTS product arrangement is VL-linker-VH, the DNA construct and the resulting product obtained by dTS is schematically represented in **Fig. 2**. In the first row, a schematic light chain RNA is depicted comprising a variable κ leader segment (Lκ) followed by an intron, a rearranged Vκ gene (V-Jκ), an intron (including the black arrow) and a constant κ segment (Cκ). In the second row, a schematic heavy chain RNA is depicted comprising a variable leader segment (LH) followed by an intron, a rearranged VH gene (V-D-JH), a CH1 domain segment, a hinge region (H) segment, and CH2 and CH3 domain segments, all separated by introns. The DNA construct of the

invention (RNA of Special Gene, third row) comprises promoter/enhancer sequences (not shown) and a template for RNA synthesis encoding an RNA transcript cpmprising: (a) a first segment (RCA of the invention), represented by the black arrow with its sense strand reverse-complementary to an adequate stretch in the light-chain (L) chain pre-mRNA of the first row: (b) a spacer sequence (SP1); (c) an exon encoding a flexible peptide linker (LINK), preferably similar to those which are widely used to create antibody scFv fragments; (d) a spacer sequence (SP2); and (e) a second segment (RCB of the invention), with reverse-complementarity to a portion of the heavy-chain (H) chain pre-mRNA of the second row. As shown in **Fig. 2**, RNA synthesized from the DNA construct of the invention will simultaneously anneal with the light chain and heavy chain RNAs, the intermediate product will be processed and the expected double trans-spliced, mature mRNA product, depicted in the last row (Final RNA Product), will in fact code for a single-chain antibody.

In another embodiment of the invention, the exon encoding the peptide linker and its flanking splice sites (LINK) is absent, and the transcript of the construct of the invention is expected to simply serve as a bridge and facilitate TS between the 3' end of the L chain V-J exon and the 5' end of the H chain V-D-J exon. The scFv encoded by such fTS will typically harbor additional four amino acids between VL and VH. This is the result of the structure of germ line Ig V genes for both H and L chains, wherein the last nucleotide in the leader exon (immediately upstream of the splice junction) contributes the first base of a codon. The last two bases of this codon are at the 5' end of the V exon. This split codon is typical to Ig, TCR and other Ig superfamily genes. Together with the following three codons (in most cases) they encode the four carboxy terminal amino acids of the leader peptide. The presence of this short peptide between VL and VH in the resulting scFvs is thus intrinsic to this design. It is important to note that in both VL-VH and VH-VL arrangements, the reading frame is not interrupted. Functional diabodies can be formed with a 5-amino acid linker and with no linker, and in the latter case even display a higher affinity for the antigen than the intact antibody (Holliger et al., 1993). This apparent peptide linker-independence is indicative of a

sufficient inherent flexibility in the corresponding termini of the joined V domains, and renders likely the assumption that fTS will result in functional structures.

The frequency of fTS or dTS events in the cell nucleus, which require two distinct encounter events mediated by the same primary transcript of the TS construct, may be hard to predict. Indeed, as a result of our design (**Fig. 2**), this transcript lacks introns. Intronless RNAs microinjected into the cell nucleus (Wang et al., 1991) reveal diffuse intranuclear patterns. Unlike, intron-containing RNAs similarly introduced into the nucleus accumulate in discrete loci, associated with spliceosome components. This distribution pattern of transcripts which lack introns is most likely due to free diffusion in the nucleus, and is the anticipated fate of the transcript of the TS construct.

Current models suggest that pre-mRNA splicing usually takes place co-transcriptionally (for review of RNA maturation and intranuclear movement, see Daneholt, 1999). In addition, inefficiently processed pre-mRNA molecules are implicated to be retained in the transcription site (Custodio et al., 1999). However, recent studies indicate that pre-mRNA molecules released from their gene template in the form of pre-messenger ribonucleoprotein (pre-mRNP), move randomly in the nucleus, apparently by free diffusion (Daneholt, 1999). As antibody H and L genes are most extensively transcribed in activated B cells during an immune response, premature release of non-spliced or only partially spliced transcripts is expected to prove sufficient for the molecular encounters required for TS. Co-localization of injected RNA with spliceosomal components (see above) suggests that transcription is not obligatory for this association. There is ample experimental evidence suggesting that spliceosome assembly and splicing are not restricted to chromosomal genes, of which the most relevant ones are those described above for in-cell spliceosome-mediated TS of extrachromosomal genes expressed transiently in the nucleus.

In the DNA construct of the invention, the L and H chains may be of any suitable vertebrate, preferably a mammalian, most preferably, a mouse or human antibody.

In one preferred embodiment of the invention, the DNA construct is designed for construction of a vertebrate's antibody library and comprises a sense strand that harbors sequences reverse-complementary to most of the vertebrate's antibody L chain transcripts and another sense strand that harbors sequences reverse-complementary to most of the vertebrate's antibody H chain transcripts.

According to this embodiment, the invention provides a DNA construct for construction of a mouse antibody library, wherein:

RCA represents a nucleotide segment of about 120 nucleotides, having one sequence reverse-complementary to the 3' region of the J-C intron of the mouse κ chain gene, said 3' region having the sequence

CTTATCTGTA GGGATAAGCG TGCTTTTTTG TGTGTTGTAT
ATAACATAAC TGTTTACACA TAATACACTG AAATGGAGCC
CTTCCTTGTT ACTTCATACC ATCCTCTGTG CTTCCTTCCT C

SP1 represents a spacer sequence;

LEX is a segment encoding a linker peptide used for construction of scFv molecules or a part of said peptide;

SP2 represents a spacer sequence; and

RCB represents a nucleotide segment containing 19 different sequences, each of which is reverse-complementary to a defined stretch of about 20 nucleotides which is centered in the leader exon/intron junction of one or more germ-line mouse VH genes.

The 19 sequences contained in RCB are reverse-complementary to the sequences 1 to 19 depicted in Fig. 9. The segment RCB is encoded by the sequence depicted in Fig. 11. The segment RCA above is encoded by the sequence between the XhoI and the NotI depicted in Fig. 12B.

A schematic structure of such a DNA construct for producing a mouse antibody library is depicted in Fig. 12, wherein RCA, SP1, LEX, SP2 and RCB correspond to the stretches Xh-N, Cgamma1-hinge intron, linker, mJH3-JH4 intron and Xb-R, respectively.

According to this same embodiment, the invention provides a DNA construct for construction of a human antibody library, wherein:

RCA represents a nucleotide segment having the following three sequences: (a) a sequence of about 120 nucleotides reverse complementary to the 3' end of the J-C intron of the human Ig κ chain gene; (b) a sequence of about 120 nucleotides reverse complementary to the 3' end of the Jλ1-Cλ1 intron of the human Ig λ1 gene; and (c) a sequence of about 50 nucleotides reverse complementary to the 3' end of the Jλ2-Cλ2 intron of the human Ig λ2 gene;

SP1 represents a spacer sequence;

LEX is a segment encoding a linker peptide used for construction of scFv molecules or a part of said peptide;

SP2 represents a spacer sequence; and

RCB represents a nucleotide segment containing 24 different sequences, each of which is reverse complementary to a defined stretch of about 20-25 nucleotides which is centered in the leader exon/intron junction of one or more germ-line mouse VH genes.

The sequences (a) to (c) of the L chain in RCA are reverse-ccomplementary to the following sequences a', b' and c', respectively:

Sequence (a'):

5' TCTGGGATAA GCATGCTGTT TTCTGTCTGT CCCTAACATG CCCTGTGATT ATCCGCAAAC AACACACCCA AGGGCAGAAC TTTGTTACTT AAACACCATC CTGTTTGCTT CTTTCCTCAG 3'

Sequence (b'):

5' CCCCGGGTGG ACCGGATGGC CACACTGTGA ACCCTCCCAG AGACTTTAGA CAGAGAGAGG GGCTCCACAA CACCCCGGTA TTCTGTCTGC CCTCTCTCAC CCCCTTCCCT GTCCACACAG 3'

Sequence (c'):

5' CCCAGGTGGA CACCAGGACT CTGACCCCCT GCCCCTCATC CACCCCGCAG 3'

This 50 bp sequence is identical in all the following introns: at the 3' end of the human Jλ2-Cλ2 intron (positions 9481-9600 in GenBank Accession X51755); Jλ3-Cλ3 intron (positions 14869-14988 in X51755); Jλ6-Cλ6 intron

(positions 28037-28156 in X51755); Jλ7-Cλ7 intron (positions 31081-31200 in X61755), and it spans the BP which is located 25 bp upstream to the acceptor site.

The 24 sequences contained in RCB are reverse-complementary to the sequences 1 to 24 depicted in Fig. 13. The segment RCB is encoded by the sequence depicted in Fig. 14. The segment RCA is encoded by the sequence depicted in Fig. 16.

A schematic structure of such a DNA construct for producing a human antibody library is depicted in Fig. 17A, wherein RCA, SP1, LEX, SP2 and RCB correspond to the stretches Xh-N, Cgamma1-hinge intron, linker, mJH3-JH4 intron and Xb-R, respectively.

In a further embodiment of the invention, the DNA construct of the invention is useful for covalent intracellular joining of selected exons from transcripts of TCR α and β chains in a T cell and either said gene A encodes a TCR α chain and said gene B encodes a TCR β chain, or said gene A encodes a TCR β chain and said gene B encodes a TCR α chain.

A single gene designed to mediate fTS or dTS of either murine or human Ig H chain and L chain mRNAs or TCR α and β chain mRNAs can be employed universally for the generation of immune antibody and TCR display libraries from different sources in a number of different modes. Most filamentous phage-display procedures and some other display technologies allow expression of more than one scFv molecule on the phage surface, and thus enable the production of either one of the two functional scFv modalities as defined herein – the one with a short linker (diabodies in the case of antibodies) and the other with the longer flexible linker.

In one embodiment of the invention, this single gene represented by the DNA construct of the invention for intracellularly joining the L and H chains of antibodies of interest, is employed as a transgene and introduced into the genome of a non-human vertebrate such as an avian, e.g. chicken, or a mammal, preferably a rodent, most preferably a mouse.

Thus, in another aspect, the invention relates to a transgenic non-human vertebrate, preferably a mammal, harboring a DNA construct of the invention wherein the genes A and B are the L and H chains of an antibody.

In one preferred embodiment, the mammal is an ordinary mouse strain routinely used for immunization and production of mAbs (Harlow and Lane, 1988). In another embodiment, the mammal is a special transgenic mouse line incapable of producing self-antibodies and harboring human Ig H and L chain gene loci. These mice, when immunized with an antigen of interest, produce the corresponding human antibodies (e.g., Mendez et al., 1997). In both cases, the transgenic mice are produced by standard procedures with the transgene of the present invention. Transspliced mRNA species are expected to be found in all B cells of these transgenic animals and to generate the scFv genes of the antibodies of interest when the mouse is immunized with the appropriate antigen.

In another embodiment, the DNA construct of the invention can be introduced by any technology for gene delivery such as retroviral infection or liposomes, for transient expression, into B and T cells isolated from antigen-exposed mice or humans or immunodeficient mice reconstituted with human lymphocytes (Lubin et al., 1991) and immunized with the antigen of choice. These techniques are particularly suitable for introducing the construct in human T cells isolated from immune individuals.

In still another embodiment, the present invention provides a library comprising a variegated population of cDNA molecules suitable for preparation of gene libraries encoding scFv molecules of antibodies of interest, said library being produced by a method comprising the steps:

(a) immunizing a transgenic mouse of the invention with an appropriate antigen to produce the antibodies of interest in said transgenic mouse;

(b) extracting RNA from mature B cells of said immunized mouse ; and

(c) subjecting said RNA preparation to reverse transcriptase reaction,

whereby a variegated population of cDNA molecules are obtained from which gene libraries encoding scFv molecules of said antibodies of interest can be produced by PCR employing variable region-specific primers.

According to the invention, RNA is extracted from mature B cells of, for example, bone marrow, peripheral blood cells (PBL), or spleen preparations, using standard protocols.

The pool of RNAs prepared from the B-cell population to which the special gene has been introduced will be amplified by RT-PCR, using a unique set of primers specific to the 5' end of VL and the 3' end of VH genes. This protocol assures that only trans-spliced mRNA are amplified. Products will then be directly cloned as scFv genes into suitable display vectors.

The present approach requires that sufficient sequence data are available both for assembly of TS constructs and synthesis of specific antibody L and H or TCR α and β chain-specific primers for RT-PCR amplification of trans-spliced RNA molecules. In the absence of a technology for the production of transgenic organisms from a given species, at least two routes can be employed for the introduction of TS-mediating nucleic acids into the nucleus of immune cells: (a) Gene delivery into the nucleus of B cells such as by retroviral infection or with liposomes; and (b) Direct introduction into immune cells of RNA molecules encoded by the TS construct, which are transcribed and preferably capped in vitro. For this purpose, one can use one of the many commercially available cloning vectors suitable for in-vitro transcription, and an in-vitro transcription (and capping) kit.

In still a further embodiment, the invention provides a gene library comprising a variegated population of double-stranded DNA molecules encoding scFv molecules of antibodies of interest, said gene library being generated by a method comprising the steps:

(a) immunizing a transgenic mouse of the invention with an appropriate antigen to produce the antibody of interest in said transgenic mouse;

(b) extracting RNA from mature B cells of said immunized mouse; and

(c) subjecting said RNA preparation to reverse transcriptase-PCR,

thus obtaining a variegated population of double-stranded DNA molecules encoding scFv molecules of said antibodies of interest.

In one embodiment, the transgenic mouse is an ordinary mouse strain and the gene library consists of DNA molecules encoding mouse scFv antibodies. In another embodiment, the mouse is a transgenic mouse invapable of producing self-antibodies and harboring human Ig H and L chain gene loci, and the gene library contains DNA encoding human scFv antibodies.

In yet still another embodiment, the invention relates to an expression vector comprising a variegated population of DNA molecules encoding scFv molecules of antibodies of interest, preferably an expression vector which is expressible and displayable on the surface of a cell or viral particle such as, but no being limited to, bacteria, phage, ribosome or yeast cells or any other display system.

In one embodiment, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein can be detected on the surface of the bacteria (Ladner et al., PCT Publication WO 88/06630). In another embodiment, the gene library is cloned into a yeast cloning vector and the scFv is displayed on yeast cells. In yet another embodiment, ribosomal display comprises the use of RNA transcribed from a DNA construct of the invention wherein said RNA is translated in a cell-free translation system.

In a most preferable embodiment, the expression vector is a phagemid suitable for expression and display in E. coli strains infected with helper phages, whereby the gene library is expressed as fusion protein on the surface of the bacteriophage. For example, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phages can be applied to affinity matrices at very high concentrations, large number of phages can be screened at one time. Second, since each infectious phage encode the scFv molecule on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be applied by another round of infection. In this embodiment, any phage used in phage display libraries such as, but not limited to, the almost identical E. coli filamentous phages M13, fd and fl, can be used in the present invention (Ladner et al., PCT Publication WO 90/02909; Garrard et al., PCT Publication WO 92/09690).

Thus, in yet another embodiment, the invention relates to a phage-display library comprising a plurality of recombinant phages each having an expression vector of the invention comprising a variegated population of DNA molecules encoding scFv molecules of antibodies of interest fused to a capsid protein of the phage such as capsid gene III or gene VIII, wherein said phage-display library is generated by co-transfecting a suitable bacterial cell with an expression vector of the invention and a helper phage. In one preferred embodiment, the bacterial cell is a E. coli strain and the helper phage is the M13 filamentous bacteriophage.

The invention further provides a method for generating an antibody library comprising a variegated population of scFv molecules of antibodies of interest expressed and displayed on the surface of a cell or viral particle, said method comprising the steps:

(a) immunizing a transgenic mouse of the invention with an appropriate antigen to produce the antibodies of interest in said transgenic mouse;

(b) extracting mRNA from mature B cells of said immunized mouse;

(c) subjecting said mRNA preparation to RT-PCR thus obtaining PCR products which consist of a variegated population of DNA molecules encoding scFv molecules of said antibodies of interest;

(d) cloning the PCR products obtained in (iii) in an expression vector; and

(e) cloning said expression vector in a suitable system whereby said DNA molecules of (c) encoding scFv molecules of antibodies of interest are expressed and displayed on the surface of a cell or viral particle.

When the library is a phage-display library, the scFv molecules of antibodies of interest are fused to a capsid protein of the bacteriophage.

The displayed scFv may be screened by panning with antigen and then isolated and converted into full antibodies by procedures well-known in the art. In this way, antibodies of a particular specificity can be selected and recovered from the phage library.

According to the approach of the present invention, immune antibody libraries from any vertebrate species can be produced. In a preferred embodiment, the vertebrate is a mammal, most preferably a mouse, that can produce either

mouse or human antibody libraries as described hereinbefore. However, avians can also be used. For example, a number of immune chicken antibody libraries have been generated in recent years for the isolation of antibodies against human antigens. The evolutionary distance between mammals and birds enables chicken to mount a strong antibody response against human antigens, which are highly conserved among mammals, and are thus only weakly immunogenic in common laboratory animals.

The present invention allows, likely for the first time, the faithful reconstitution of entire antibody immune repertoires in-vitro as libraries of scFvs displayed on phage or other display means. Such a capacity should combine the extraordinary ability of the immune system to produce specific, high affinity antibodies in response to antigen, with the fast and easy protocols of in-vitro display technologies.

The invention further provides a method for generating a variegated population of cDNA molecules suitable for preparation of gene libraries encoding scFv molecules of human TCRs of interest, said method comprising the steps:

(a) introducing a DNA construct of the invention wherein either gene A encodes a TCR α chain and gene B encodes a TCR β chain, or gene A encodes a TCR β chain and gene B encodes a TCR α chain, to T cells isolated from an immune individual;

(b) extracting RNA from the T cells of (a); and

(c) subjecting said RNA preparation to reverse transcriptase reaction,

whereby a variegated population of cDNA molecules are obtained from which gene libraries encoding scFv molecules of said TCRs of interest can be produced by PCR employing variable region-specific primers.

The gene libraries obtained from these cDNA libraries by PCR can then be cloned into an expression vector such as an expression vector which is expressible and displayable on the surface of a cell or viral particle such as, but not being limited to, a phagemid.

The invention further provides a phage-display library comprising a plurality of recombinant phages each having an expression vector of the invention, wherein

said expresssion vector comprises DNA molecules encoding scFv molecules of human TCRs fused to a capsid protein of the phage, such library being generated by co-transfection of a suitable bacterial cell with said expression vector and a helper phage. The bacterial cell is preferably E. coli and said helper phage is for instance M13 filamentous bacteriophage.

There is further provided a method for generating a human TCR library comprising a variegated population of scFv molecules of human TCRs of interest expressed and displayed on the surface of a cell or viral particle, said method comprising the steps:

(a) introducing a DNA construct as described above to T cells isolated from an immune individual;

(b) extracting RNA from the T cells of (a);

(c) subjecting said mRNA preparation to RT-PCR, thus obtaining PCR products which consist of a variegated population of DNA molecules encoding scFv molecules of said human TCRs of interest;

(d) cloning the PCR products obtained in (c) in an expression vector; and

(e) cloning said expression vector in a suitable system whereby said DNA molecules of (c) encoding scFv molecules of human TCRs of interest are expressed and displayed on the surface of a cell or viral particle.

When said library is a phage-display library, said scFv molecules of human TCRs of interest are fused to a capsid protein of the phage.

The invention will now be illustrated by the following non-limiting Examples.

EXAMPLES

Materials and Methods

(a) Cells

SP6 is a mouse B cell hybridoma, producing an IgM(κ) anti-2,4,6-trinitrophenyl (TNP) antibody (Kohler and Shulman, 1980), kindly provided by Prof. Z Eshhar, The Weizmann Institute of Science, Rehovot, Israel.

COS7 cells are transformed African Green Monkey kidney fibroblasts (Gluzman, 1981).

Cells were cultured in DMEM, supplemented with 10% heat-inactivated fetal serum (HI-FCS), 2 mM L-Glutamine, 1 mM sodium pyruvate, Pen-Strep Solution (penicillin 10,000 units/ml, streptomycin, 10 mg/ml at 1:1000 dilution), all from Biological Industries (Beit Haemek, Israel), in the presence of 7.5% CO_2, at 37°C.

(b) Plasmids

pR-Tκ1 is the SP6 genomic κ chain gene and pR-SP6 is the SP6 genomic H chain gene (Ochi et al., 1983). Both were a gift from Prof. Zelig Eshhar, The Weizmann Institute of Science, Rehovot, Israel.

pSV-Vμ1 (Neuberger, 1983) is a genomic clone containing a rearranged mouse VH gene, JH segments, the JH-Cμ intron and the full mouse Cμ gene. B-102 is a genomic clone encoding a full mouse κ chain. Both were a gift from Dr. M.S. Neuberger, Cambridge, UK.

pNeoγ1 is a plasmid containing the full length genomic human Cγ1 gene, and pgpt-VL1 is a plasmid containing the full length genomic human Cκ gene. Both were a gift from Dr. S. Dagan, XTL, Rehovot, Israel.

pBJ1-Neo, a mammalian expression vector (Lin et al., 1990), was kindly provided by Prof. Zelig Eshhar, The Weizmann Institute of Science, Rehovot, Israel.

pBlueScript II KS(-), a commercial cloning vector (Stratagene La Jolla, CA, USA)

pGEMT is a commercial cloning vector for PCR products (Promega Co., Madison, WI, USA).

(c) SP6 hybridoma cells as a model system

In the experimental systems used herein in the examples, it was decided to incorporate a peptide linker exon, so as to allow the expected dTS processes to

occur within the nuclei of transfected cells. As all elements required for fTS are included in our final constructs, products lacking the linker sequence will still allow analysis of the fTS events and characterization of their expressed polypeptide products. In order to be able to define and characterize genetic components required for both sTS and dTS processes, it was decided first to focus on a pair of H and L chain genes encoding a well-characterized antibody of a known specificity. For this purpose, we chose the SP6 mouse B-cell hybridoma that produces an anti-trinitrophenyl (TNP) mAb (Kohler and Shulman, 1980). The full nucleotide sequence of this mAb genes is available at the GenBank Accession No. J00569 for the light chain and No. J56936 for the heavy chain.

(d) Stable Transfection of SP6 Cells by Electroporation

$5x10^6$ SP6 hybridoma cells were harvested and resuspended in 800 µl of growth medium, and 15 µg of linearized plasmid DNA were added. Electroporation was performed by the EasyJect Plus apparatus (EquiBio Ltd., Ashford, UK) in a 4 mm cuvette (Voltage: 200 V; Capacity: 750 mF; Resistance: infinite), at room temperature. Cells were immediately diluted in growth medium and divided into 96-well plates, at 100 µl per well. Two days later, G418 (G418 Sulphate, Life Technologies Glasgow, UK) was added to a final concentration of 1 mg/ml, and a final volume of 200 µl per well. G418-resistant clones were detected and expanded 1-2 weeks post-transfection.

(e) Transient Transfection of COS7 Cells by the DEAE-Dextran Method

COS7 cells were grown in 90 mm plates to 80% confluence, and washed once with fresh PBS buffer (1.8 mM KH_2PO4, 10 mM Na_2HPO_4, 140 mM NaCl, 2.7 mM KCl). Then 5 ml of Solution A (0.1 M Tris-HCl (pH 7.2), 0.5 mg/ml DEAE-Dextran (Amersham Pharmacia, Piscataway, NJ, USA),) containing 3 µg of plasmid DNA were added to the cells. After 30 minutes of incubation at 37°C, 5 ml of Solution B (growth medium containing 200 µM chloroquine (Sigma St Louis, MO, USA,) were added, and cells were incubated for further 2.5 hours at 37°C.

After removal of the supernatant, cells were incubated in the presence of 5 ml Solution C (growth medium containing 10% DMSO) for 2.5 minutes. Solution C was removed and 12 ml of growth medium were added. Three days post-transfection cells were harvested for mRNA isolation and growth medium was collected for ELISA.

(f) Molecular Biology Techniques

All standard cloning and gene manipulation techniques were performed according to Sambrook et al., 1989.

Restriction enzymes and DNA-modifying enzymes were from New England Biolabs (Beverly, MA, USA)

mRNA was isolated from $5x10^6$ SP6 or COS7 cells using the Oligotex Direct mRNA Mini Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions.

PCR analysis was performed with the HotStarTaq Master Mix Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions.

RT-PCR was performed with the QIAGEN One Step RT-PCR Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions.

Oligonucleotides were synthesized by Bio-Technology General, Rehovot, Israel.

PCR and RT-PCR products were cloned directly into the pGEM-T Vector System according to the manufacturer's instructions.

(g) Construction of Synthetic DNA Fragments

The longest oligonucleotide was suspended in water to a final concentration of 1 μg/μl. All other oligonucleotides were suspended to a similar molar concentration as the longest one. For the assembly of the DNA fragment, all the oligonucleotides (1 μl each) were mixed, boiled for 5 minutes, cooled at room temperature, and ligated overnight with T4 DNA Ligase, at 16°C. The resulting fragment was cloned into pBlueScript II KS(-) or pBJ1-Neo, via the cohesive ends which were created at the ends of the assembled synthetic DNA fragment. DNA sequence of insert was

determined. Unique restriction sites incorporated into the fragment were used for complementation of mutations introduced during the cloning procedure.

(h) ELISA

All reagents were from Sigma (St Louis, MO, USA). The assay was performed in triplicates in a microtiter plate. Incubations were done at 37°C for 1 hour in 100 µl, except blocking and buffer-substrate solutions, of which 200 µl was added. 3 washes were done between the different steps with phosphate-buffered saline (PBS) containing 0.05% Tween 20. Wells were coated with 5 µg/ml polyclonal goat anti-human IgG antibody (Fc-specific) diluted in PBS, and blocked with PBS containing 2% of bovine serum albumin (BSA). Several dilutions of samples to be tested or commercial human IgG1 antibody were then added. The assay was developed with polyclonal goat anti-human IgG (Fc-specific)-horseradish peroxidase-conjugated (at 1:20,000 dilution) and OPD substrate-buffer tablets dissolved in double distilled water. Optical density at 450 nM was read with SLT Spectra ELISA reader (SLT-Labinstruments GmbH, Salzburg, Austria).

(i) The Arrangement of VH and VL in the TS Product scFv

An antibody combining site can often be reconstituted by the Fv portion of the antibody through the addition of a synthetic peptide linker, expressed as a scFv. Both possible arrangements of the VH and VL regions, namely (going from the N to the C terminus) either VL-linker-VH or VH-linker-VL, can very often regenerate the original antibody-binding profile. Because of significant differences in number and complexity of genetic elements that will have to be incorporated into the final dTS constructs (see below), the VL-linker-VH construct was chosen for the present experiments, as illustrated in **Fig. 2**. However, constructs based on VH-linker-VL arrangements are similar in essence and are feasible as well.

The chosen arrangement can be obtained by dTS only if the 5' part of the transcript of the dTS construct contains sequences reverse complementary to the JL-CL intron, while its 3' part harbors a stretch reverse complementary to a region upstream of VH genes, namely, the leader or leader-VH intron (see **Fig. 2**).

(j) The Peptide Linker in the dTS Produced scFv

The TS RNA products have been designed to encode the SP6 antigen-binding site as a scFv. In previous work (Gross and Eshhar, 1992), we have already shown that SP6 VH and VL chains can form a functional scFv. The peptide linker chosen for the present experiments to join SP6 VH to VL domains is a 18-amino acid long peptide described before designated 218 (Whitlow and Filpula, 1992) and having the sequence:

Gly-Ser-Thr-Ser-Gly-Ser-Gly-Lys-Pro-Gly-Ser-Gly-Glu-Gly-Ser-Thr-Lys-Gly

It is encoded by the nucleotide sequence:

5' GGC TCT ACT TCC GGT TCA GGA AAG CCC GGA TCC GGT GAA GGT AGC ACT AAA GGT 3'

The underlined BamHI restriction site was introduced into the middle of the sequence coding for this peptide linker (see sequences of PCR primers below) to enable cloning, replacing an original SmaI site without changing the amino-acid composition of the linker.

Example 1

DNA construct for sTS of SP6 L chain RNA + linker with SP6 H chain RNA

In order both to dissect dTS into its two sTS components, and to better assess TS of transcripts of two chromosomal genes in the cell nucleus, we first assembled a genetic construct designed to join the linker exon (see **Fig. 2**) with SP6 VH exon, so as to result in a full scFv-encoding mRNA.

In order to achieve our goal, the following issues were taken into consideration: This construct should harbor a spacer sequence between the splice donor site immediately downstream to the linker exon and the stretch reverse-complementary to the leader sequence (see **Fig. 3**). Without the spacer sequence, the donor splice junction flanking the linker exon may be inaccessible to the spliceosome as a result of the local bi-molecular RNA structure. The sequence chosen for this function is derived from the intron between the mouse Cμ4 exon

(coding for the 4th domain of mouse IgM constant region) and the Cμ membranal exon.

Trans-splicing of the peptide linker exon to the VH exon would result in extra 4 codons, which would be added to the linker 218 of (j) above. The first option was to allow this situation, adding a tetrapeptide of the sequence Gly-Val-His-Cys to the peptide linker, thus creating amino acid positions 19-22. The second option was to genetically cut the peptide linker 4 codons shorter in order to add the tetrapeptide while keeping unchanged the final peptide linker size (18 amino acids). In this case, the peptide linker composition in positions 15-18 would change from Ser-Thr-Lys-Gly to Gly-Val-His-Cys. Linkers of 15-amino acids in the form of $(Gly_4Ser)_3$ are commonly used for the production of scFv fragments, and the Gly at position 15 complies with their amino acid composition. The effect of the amino acids Val-His-Cys at the C terminus of the linker, which do not usually form a part of the scFv peptide linkers on the scFv structure, remains to be seen.

The second option was chosen and the following clone was assembled from two fragments:

1. A 0.65 kb XhoI/BamHI PCR fragment containing the rearranged SP6 Vκ gene (including the leader and L-V intron) and the 5' half of the sequence encoding the synthetic linker, including the BamHI site, joined in-frame. It was amplified from clone pR-Tκ1 (Ochi et al., 1983) with the following primers:

The 5' primer, 16919, contains an XhoI site and the ATG codon, and covers positions 174 to 191 in the gene (GenBank Accession No. J00569):

5' GCG CTC GAG CAG ACC AGC ATG GGC TTC 3'

The 3' primer, 5749 contains the 3' end of VJκ (positions 732-752 in GenBank Accession No. J00569), joined in frame to the 5' part of the linker and the BamHI site:

5' CGC GGA TCC GGG CTT TCC TGA ACC GGA AGT AGA GCC TTT CAG CTC CAG CTT GGT CCC 3'

2. An 0.35 kb BamHI/NotI fragment containing the 3' part of linker 218 (from the BamHI site) down to position 14 (see (c) above)with an extra G (to be spliced to

SP6 VH exon), a consensus donor splice site, a stretch from the Cμ intron (positions 241-559 in GenBank Accession No. K01238) as a spacer and a sequence (60 bp) reverse-complementary to most of the SP6 VH leader exon and the 3' part of the V-L intron, covering all the donor splice junction (positions 794 to 853 in GenBank Accession No. X56936). This configuration should create a stable double-stranded RNA structure, which is expected to render this donor splice site non-functional. The template used for PCR amplification of this fragment was the plasmid pSV-Vμ1 (Neuberger, 1983) containing the full genomic mouse Cμ gene.

The 5' primer, 4396, contains 3b protection, the 3' part of the linker including the BamHI site, and the 5' end of the Cμ stretch:

5' GCG GGA TCC GGT GAA GGA GGT AAG TGC TCT GCA CAC ACC CTG C 3'

The 3' primer was 8911. In the original orientation, it contains the 3' end of the Cμ sequence, the 60b from SP6 LVH and L-V intron in areverse-complementary configuration, a NotI site and 3b protection. In the final primer sequence, the SP6-derived stretch is therefore identical to its SP6 VH origin:

5' CGC GCG GCC GCT GGA GCC GGA TCT TTC TCT TCC TCC TGT CAA TAA TTG CAG GTA AGG GGC TCA CCA ATT CCC CAT GCA TGT GTG TGC GC 3'

The two fragments were cloned in a single step into the pBJ1-Neo mammalian expression vector (Lin et al., 1990) using XhoI + NotI, to produce clone 965-14, which is presented in **Fig. 3**.

Example 2

sTS demonstration in SP6 hybridoma cells

In this experiment, the capacity of the transcript of clone 965-14 of Example 1 to mediate TS with the endogenous SP6 H chain transcript, inside the nucleus of stably-transfected SP6 hybridoma cells, was explored. TS of 965-14 RNA with SP6 H chain transcript should yield a scFv-encoding segment attached to the mouse Cμ (the SP6 isotype).

Plasmid 965-14 was transfected into SP6 cells by electroporation, and G418-resistant clones were selected and expanded. Messenger RNA was prepared from several stable transfectants, and was analyzed by RT-PCR for expression of the introduced DNA, along with expression of endogenous SP6 VH or VL genes. Transfectant 965-2, which was positive for expression of the transfected DNA, was further analyzed for sTS by a second round PCR, using a nested set of primers. PCR products were cloned into PCR cloning vector pGEMT (Promega) and the DNA sequence of two clones (979-1 and 979-3) was determined. These sequences, shown in **Fig. 4**, reveal a precise sTS product. The G nucleotide at the 3' end of the linker was spliced to the 5' end of SP6 VH exon, creating, together with the GT dinucleotide, a GGT codon (encoding Gly), thus preserving the reading frame of the scFv gene.

As mRNA of clone 965-2 was prepared several weeks post-transfection, the introduced plasmid DNA must have been present in the nucleus of these cells only as stably-integrated chromosomal DNA. Hence, this is not only a clearcut demonstration of intracellular joining of a peptide linker to an antibody V gene via sTS, but also probably the first example of TS of exons from transcripts of two different chromosomal genes. This finding supports our assumption that, although splicing is believed to be generally carried out co-transcriptionally, TS between the transcripts of two different genes, which most likely reside at unrelated chromosomal locations, can still occur.

Example 3

Constructs for TS in COS7 cells

The main advantages offered by the SP6 hybridoma are that it is an antibody-producing B cell synthesizing high levels of H and L chain transcripts and that stable transfectants expressing an introduced gene can be isolated and easily propagated. However, obtaining stable transfectants with the desired phenotype is a time-consuming procedure. Hence, analysis of a series of modifications in a tested genetic construct is a lengthy and laborious process. Moreover, as a hybridoma, it is a cell with an unknown chromosomal content and may, therefore, display an

abnormal pattern of certain transcription factors, a shortcoming likely to have an effect on expression of introduced genes. In addition, the SP6 hybridoma may express a very high level of a transcript from a non-productively rearranged L chain allele of its NSO fusion partner, as shown for all myeloma fusion partners tested for its presence (Carroll et al., 1988).

In order to circumvent these difficulties, we have chosen to employ the COS7 transient expression system (Sambrook et al., 1989), and combine it with SP6 VH and VL genes.

COS7 cells provide both a powerful tool for rapid evaluation of genetic constructs (days compared with weeks to a few months) and exceptionally high level of expression of genes driven by the pBJ1-Neo strong SRα promoter. This vector also contains SV40 origin of replication, required for T-antigen binding and unchecked plasmid replication (Sambrook et al., 1989). However, since COS7 cells do not express any endogenous antibody genes, these genes have to be co-introduced into the system together with the dTS construct.

Thus, employing the COS7 transient expression system for studying dTS and its underlying two sTS processes separately with the SP6 genes, requires five different genetic constructs as defined in Examples 3a - 3e below. Several modifications were made in the genetic elements incorporated in the constructs, which are to drive TS reactions.

3a) SP6 H chain

This construct, 75-2, is depicted in **Fig. 5** and encodes a full IgG1 heavy chain with the rearranged SP6 VH. It was prepared as follows:

1. A 1.8 kb NotI/HindIII PCR product harboring the full human Cγ1 gene (GenBank Accession No. Z17370), including the 3' part of its upstream intron, was generated from a plasmid with the following primers:

The 5' primer, 24884, spans positions 51-68 of the gene and carries a NotI site:

5' GCG GCC GCT AAG GTG AGG CAG GTG 3'

The 3' primer, 23671, goes from positions 1810 to 1827 of the gene and contains the stop codon (= 3' end of Cγ3) and a HindIII site:

5' GCG AAG CTT GCC GGC CGT CGC AC 3'

This fragment was inserted into the NotI-HindIII sites at the polylinker of vector pBJ1-Neo to create cassette 56-1.

2. The rearranged genomic SP6 VH gene, with its leader, was produced as a 0.55 kb XhoI/NotI PCR fragment from the genomic SP6 heavy chain clone pR-SP6 (Ochi et al., 1983), using the following primers:

The 5' primer, 16630, harbors an XhoI site and spans positions 788-805 in the gene (Genbank Accession X56936), including the ATG codon:

5' CTC GAG ATG GGA TGG AGC CGG ATC 3'

The 3' primer, 16631, contains a NotI site, and goes from position 1318 to 1335, a sequence located within the J-C intron in the gene, approximately 40 bp downstream to the donor splice site of the J segment (mouse JH3 in this case):

5' GCG GCC GCT CAG AAT CCC CCC AAC 3'

This XhoI/NotI PCR product was inserted into cassette 56-1, to yield clone 75-2.

3b) SP6 L chain

A 1.5 kb XbaI/BamHI fragment containing genomic human Cκ from plasmid pgpt-VL1 was inserted into the multiple cloning site of pBluescript II KS (-) vector (= KS, Stratagene), excised as a NotI/EcoRI fragment and inserted into pBJ1-Neo, to produce cassette 62-1. The rearranged SP6 Vκ gene cloned in pR-Tκ1 of Example 1 above was inserted into this cassette as a 0.65 kb XhoI/NotI fragment, yielding clone 76-1.

The primers used for this cloning were the 5' primer 16919 (see Example 1 above) and the 3' primer 16633:

5' GCG GCC GCG TGT ACT TAC GTT TCA GCT CC 3', which harbors a NotI site and corresponds to the 3' end of SP6 Jκ (positions 743-762 in Genbank Accession No. J00569), including the donor splice site.

Since our TS design exploits mouse Jκ-Cκ intronic sequences and clone 76-1 harbors the human intron, a new clone, 1083-1 (depicted in **Fig. 6**), was therefore assembled, replacing the relevant stretch of the human intron in clone 76-1 with the mouse one. Plasmid 76-1 (9 kb) was digested with BamHI and XbaI, and the 7 kb fragment was isolated. Plasmid B-102 was digested with BamHI, XbaI, and BglI, and the 3.8 kb BamHI/XbaI fragment was isolated and ligated with the 7 kb BamHI/XbaI fragment from plasmid 76-1.

3c) New gene for sTS of SP6 L chain RNA + linker with SP6 H chain RNA

Two modifications have been introduced into the first version of the gene/clone 965-14 described in Example 1 above. First, the spacer region was derived from a different source so as to be free of any potential splice sites, even of low strength, which may interfere with the desired TS reaction. Second, reverse complementarity with the SP6 VH upstream sequence was reduced to 20 nucleotides, so as to test the capacity of this minimal length to mediate TS. This has strong implications to the design of the library constructs as described hereinafter. This new clone, 1033-1 (depicted in **Fig. 5**), consists of two fragments:

1. The 0.65 kb XhoI/BamHI fragment of clone 965-14 (see Example 1 above).
2. An 0.2 kb BamHI/EcoRI fragment encoding the 3' end of linker 218 as described for clone 965-14 in Example 1, followed by a donor splice site, a 150 bp spacer region from the mouse JH3-JH4 intron (nucleotides 1181-1330 in GenBank Accession No. V00777], devoid of any donor or acceptor splice sites, and a 20 bp sequence with a sense strand reverse complementary to the 3' end of the SP6 VH leader coding region, and its downstream donor splice site (nucleotides 825-844 in Genbank Accession X56936), so as to reduce competing cis-splicing events. This 20 bp stretch is flanked by XbaI and EcoRI, introduced with the 3' primer used to synthesize this fragment.

The 5' primer was 5079, with a BamHI site:

5' CCC GGA TCC GGT GAA GGA GGT AAG TTG CAC AGG CAG GGA ACA GAA TGT GG 3'

The 3' primer was 6094, with an EcoRI site:

5' CCC GAA TTC TAA TTG CAG GTA AGG GGC TCT CTA GAC AAT AGT GGG TTT TTC CTC TGT ACC 3'

3d) New gene for sTS of SP6 H chain RNA + linker with SP6 L chain RNA

This clone, 1026-2 (depicted in **Fig. 6**) comprises the following components:

1. A 120 bp XhoI/NotI fragment with a sense strand reverse complementary to the 3' region of the mouse Jκ-Cκ intron, predicted to contain the branch point (BP) sequence and including the full acceptor splice site. This 3' region (positions 4493-4613 in GenBank Accession No. V00777) has the sequence:

CTTATCTGTA GGGATAAGCG TGCTTTTTTG TGTGTTGTAT
ATAACATAAC TGTTTACACA TAATACACTG AAATGGAGCC
CTTCCTTGTT ACTTCATACC ATCCTCTGTG CTTCCTTCCT C

This sequence in the mRNA transcript is designed to anneal to the corresponding sequence in the mouse Jκ-Cκ intron and to reduce cis-splicing reactions. This fragment was amplified using the following primers:

The 5' primer, 271109 harbors an XhoI site:

5' GCG CTC GAG GAA GGA AGC ACA GAG GAT 3'

The 3' primer, 31550, contains a NotI site:

5' CGC GCG GCC GCT TAT CTG TAG GGA TAA GCG T 3'

2. An 0.4 kb NotI/BamHI fragment with an intron sequence, containing a BP and an acceptor splice site, followed by the 5' half of the sequence coding for linker 218, including the BamHI site. The intron chosen was the human Cγ1-hinge intron, and the template was a plasmid containing the genomic human Cγ gene.

The 5' primer was 25657, containing a NotI site and encompassing positions 514-531 in the gene (Medline Accession Z17370):

5' GCG GCC GCC AGC ACA GGG AGG GAG G 3'

The design of the 3' primer took into account the fact that the first nucleotide in the first codon of the linker peptide is contributed by the Vκ exon, and is a C. Therefore, the first triplet cannot encode glycine. From the amino acids encoded by triplets which start with a C, the choice was glutamine as a polar residue, encoded

by a CAA triplet, and it replaces glycine in the first position of the expected TS protein product. The primer, 5462, harbors positions 878-896 in the Cγ gene (including a PstI site at the acceptor site) and the 5' part of the linker sequence, including the BamHI site:

5' CGC GGA TCC GGG CTT TCC TGA ACC GGA AGT AGA TTC TGC AGA GAG AAG ATT GGG 3'

3. A 2.2 kb BamHI/HindIII PCR fragment coding for the 3' part of the linker peptide, joined in-frame to the 5' end of the gene for the SP6 VH polypeptide, followed by part of the intron upstream to the human Cγ1 and the full Cγ1. The template for this PCR reaction was plasmid 75-2 (see Example 3a above).

The 5' primer, 4881, encodes the 3' part of the linker, including the BamHI site, joined in-frame to the 5' end of the mature SP6 VH (from position 928 in GenBank Accession No. X56936):

5' GCG GGA TCC GGT GAA GGT AGC ACT AAA GGT CAG GTC CAG CTG CAG CAG 3'

The 3' primer, 23671, is described above in Example 3a.

3e) Gene for dTS of SP6 H chain RNA with SP6 L chain RNA

This dTS construct, clone 1034-2, is, in fact, composed of two fragments derived from the clones described above in 3c and 3d :

1.The 0.55 kb XhoI/BamHI from clone 1026-2 (described in 3d above; **Fig. 6**), which directs sTS to the κ chain transcript.

2. The 0.2 kb BamHI/EcoRI fragment from clone 1033-1 (described in 3c above; **Fig. 5**), directing TS to the H chain transcript.

The two fragments were cloned in one step into pBJ1-Neo, to produce clone 1034-2. It is presented schematically in **Fig. 7A**, and its complete nucleotide sequence is shown in **Fig. 7B**.

Example 4

Transfection of the constructs into COS7 cells

In order to evaluate sTS and dTS mediated by the constructs of Example 3 above, a series of six transfections into COS7 cells was performed, using the DEAE-dextran method (see "Materials and Methods"). Transfections 1 and 2 served as positive controls for the transfection procedure, as the efficiency can easily be monitored by the amount of human IgG1 secreted. Transfection 3, with vector DNA only, served as a negative control. Transfections 4 and 5 were designed to evaluate both reciprocal sTS processes, and transfection 6 aimed at detecting dTS events. The transfection efficiency was evaluated by ELISA performed 72 hours post-transfection, designed to detect human IgG1 secreted into the growth medium. The results are shown in the Table below, as concentrations of human IgG in the respective transfections, calculated according to a commercial human IgG1 standard curve.

Table. Production of human IgG in transfected COS7 cells

No.	Transfected clones	μg/ml human IgG	Description
1	75-2 + 76-1	3.0	VH + Vκ
2	75-2 + 1083-1	17.2	VH + Vκ
3	pBJ1-Neo	0	Null control
4	75-2 + 1033-1	2.1	sTS of linker to VH
5	1083-1 + 1026-2	0.14	sTS of Vκ to linker
6	75-2 + 1083-1 + 1034-2	14.3	dTS

Note the presence of human IgG in the supernatant of transfection 5, which is the likely protein product of accurate sTS in this transfection (as referred to in the Discussion hereinafter).

Example 5

RT-PCR analysis for the presence of sTS and dTS product mRNA

Transfected COS7 cells of Example 4 were harvested 72 hours post-transfection, and polyA RNA was extracted. Two hundred ng polyA RNA from each transfection served as a template for RT-PCR analysis (QIAGEN OneStep